

**A NOTE ON ESTIMATING THE NUMBER OF SUPER IMPOSED
EXPONENTIAL SIGNALS BY THE CROSS-VALIDATION APPROACH**

Y. Wu, K.W. Tam, F. Li and M.M. Zen

Technical Report 99-13

August 1999

Center for Multivariate Analysis
417 Thomas Building
Penn State University
University Park, PA 16802

Research work of authors was partially supported by the Army Research Office under Grant DAAHO4-96-1-0082. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

DTIC QUALITY INSPECTED 4

19991101 115

A note on estimating the number of super imposed exponential signals by the cross-validation approach

Y. Wu¹

York University, Toronto, Ontario, Canada

K. W. Tam and F. Li

Portland State University, Portland, Oregon, USA

M. M. Zen

National Cheng-Kung University, Tainan, Taiwan

Abstract

In this paper, a procedure based on the delete-1 cross-validation is given for estimating the number of super imposed exponential signals, its limiting behavior is explored and it is shown that the probability of overestimating the true number of signals is larger than a positive constant for sample size large enough. Also a general procedure based on the cross-validation is presented when the deletion precedes according to a collection of subsets of indices. The result is similar to the delete-1 cross-validation if the number of deletion is fixed. The simulation results are provided for the performance of the procedure when the collections of subsets of indices are chosen as those suggested by [7] in a linear model selection problem.

¹Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

1. INTRODUCTION

Detection of signals transmitted in the presence of noise arises in communications, radio location of objects, seismic signal processing and computer-assisted medical diagnosis, etc. Many procedures for determining the number of signals can be found in the literature. Generally speaking, except those based on the Bayesian approach, they may be classified into three types. The procedures of the first type are based on the information theory which have two components: one measures the goodness of fit and another penalizes the overestimation, such as AIC (see [2]). The procedures of the second type aim to decide the number of signals by minimizing the description length, which was explained in [5, 6]. The procedures of the third type are based on the idea of minimizing the prediction error; the use of the cross-validation approach is among them. There is no definite classification of these procedures. Some are proposed based on information theory, but they may also be counted as or are asymptotically equivalent to procedures of the second type, and vice versa, e.g. MDL. The same comments apply to the relationships between the procedures of the second and third types or between the procedures of the first and third types. Among all procedures, those based on the cross-validation approaches are intuitively attractive. In this paper, we shall discuss the performance of a procedure which decides the number of super imposed exponential signals by the cross-validation approach.

The linear model selection problem and the problem of determination of the number of signals are closely related. Recently there appeared some papers on the cross-validation approach in a linear model selection problem. The main reason for the activity now is that the heavy computation required by the cross-validation approach is no more an obstacle in today's modern computing era. At this point it is worthy to describe the contributions made by Prof. Shao on this subject. In [7], the performance of the cross-validation approach in a linear model selection problem was discussed, where it was shown that the procedure was not consistent if a fixed number of observations was

deleted for the purpose of the cross-validation and it was discovered that instead of deleting a fixed number of observations, if the number of deletion tended to infinity at certain rate, the resulting procedure would be consistent. The result was justified theoretically and its good small sample performance was supported by simulation. [8] further discussed the case that the number of regressors is not finite. The linear model selection problem is simpler than the nonlinear one we consider here. It is important to explore the performance of the cross-validation approach in the determination of the number of super imposed exponential signals with a number of deletion fixed or varied with the sample size.

Consider the following undamped exponential signal model

$$y(n) = \sum_{j=1}^{p_0} \alpha_j e^{i\omega_j n} + w(n), \quad n = 1, \dots, N, \quad (1.1)$$

where $i = \sqrt{-1}$, $\{\alpha_j\}$ is a set of unknown complex amplitudes, $\{\omega_j\}$ is a set of unknown angular frequencies, $\{w(n)\}$ is a sequence of independently and identically distributed complex random noise variables, and p_0 is unknown. Hereon, it is assumed that p_0 is bounded by some known finite number P . With this model there are two problems, the determination of p_0 and the estimation of the parameters α_j 's and ω_j 's. It was suggested in [4] to use the cross validation to estimate the p_0 when the sample size is relatively small. To obtain the cross-validation error for a particular number of signals p , it is required to estimate efficiently α_j 's and ω_j 's in presence of the missing observations. Such a method can be found in [3] where an efficient procedure for estimating the non-linear parameters, namely ω_j 's, is provided when one or more observations are missing. Note that once an efficient estimates of the non-linear parameters are found, the estimation of the linear parameters can be obtained by simple linear regression. [1] is another reference on this subject.

In this paper, the details of the procedure based on the delete-1 cross-validation is given in Section 2, a general procedure based on the cross-validation with the number of deletion fixed or varied with the sample size will be presented in Section 3, and the

simulation results are provided in Section 4 along with the discussion.

2. THE DELETE-1 CROSS-VALIDATION

Throughout the paper, it is assumed that

$$E(w(1)) = 0, \quad E(w(1) \overline{w(1)}) = \sigma^2, \quad E|w(1)|^4 < \infty, \quad (2.1)$$

σ^2 is unknown, and the true number of super imposed exponential signals is p_0 .

Denote for $p \leq P$,

$$Y = \begin{bmatrix} y(p+1) & y(p) & \cdots & y(1) \\ y(p+2) & y(p+1) & \cdots & y(2) \\ \cdots & \cdots & \cdots & \cdots \\ y(N) & y(N-1) & \cdots & y(N-p) \end{bmatrix} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N-p}]',$$

and $\mathbf{y}_{(N)} = (y(1), \dots, y(N))'$. Let $A_{(-n, N, p)}$ be the matrix obtained from the matrix Y with the rows having the observation $y(n)$ removed, $\theta_{(-n, N, p)} = \{n' : (y(n' + p), \dots, y(n')) \text{ is a row of } A_{(-n, N, p)}\}$, and $r_{(-n, N, p)}$ be the number of rows in the matrix $A_{(-n, N, p)}$. For convenience, we will omit N in the above notations.

Write $\hat{\Gamma}_{(-n, p)} = r_{(-n, p)}^{-1} A_{(-n, p)}^* A_{(-n, p)} = (\hat{\gamma}_{m\ell}^{(-n, p)})$, and let $\hat{\lambda}_1^{(-n, p)} \geq \dots \geq \hat{\lambda}_{p+1}^{(-n, p)}$ be the eigenvalues of $\hat{\Gamma}_{(-n, p)}$ and $\hat{\mathbf{b}}^{(-n, p)} = (\hat{b}_0^{(-n, p)}, \dots, \hat{b}_p^{(-n, p)})'$ be a unit eigenvector of $\hat{\Gamma}_{(-n, p)}$ corresponding to $\hat{\lambda}_{p+1}^{(-n, p)}$. Denote the solutions of $\sum_{j=0}^p \hat{b}_j^{(-n, p)} z^j = 0$ by $\hat{\rho}_j^{(-n, p)} e^{-i\hat{\omega}_j^{(-n, p)}}$, $j = 1, \dots, p$, where $\hat{\rho}_j^{(-n, p)} > 0$. Define

$$\hat{X}_{(-n, p)} = \begin{bmatrix} e^{i\hat{\omega}_1^{(-n, p)}} & \cdots & e^{i\hat{\omega}_p^{(-n, p)}} \\ \cdots & \cdots & \cdots \\ e^{iN\hat{\omega}_1^{(-n, p)}} & \cdots & e^{iN\hat{\omega}_p^{(-n, p)}} \end{bmatrix}.$$

Let $\tilde{X}_{(-n, p)}$ be a matrix with the n th row removed from $\hat{X}_{(-n, p)}$, and

$$\mathbf{y}_{(-n, N)} = (y(1), \dots, y(n-1), y(n+1), \dots, y(N))'.$$

Denote the least square estimates of $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ using samples $\mathbf{y}_{(-n, N)}$ and $\mathbf{y}_{(N)}$ by

$$\hat{\boldsymbol{\alpha}}_{(-n, p)} = (\hat{X}_{(-n, p)}^* \tilde{X}_{(-n, p)})^{-1} \hat{X}_{(-n, p)}^* \mathbf{y}_{(N)},$$

$$\tilde{\alpha}_{(-n,p)} = (\tilde{X}_{(-n,p)}^* \tilde{X}_{(-n,p)})^{-1} \tilde{X}_{(-n,p)}^* \mathbf{y}_{(-n,N)}, \quad (2.2)$$

respectively and write $\hat{\mathbf{x}}_{(n,p)} = (e^{in\hat{\omega}_1^{(-n,p)}}, \dots, e^{in\hat{\omega}_p^{(-n,p)}})$. Let

$$CR(p, N) = (1/N) \sum_{n=1}^N \|y(n) - \hat{\mathbf{x}}_{(n,p)}^* \tilde{\alpha}_{(-n,p)}\|^2,$$

for $p = 1, \dots, P$. If $CR(\hat{p}, N) = \min_{1 \leq p \leq P} CR(p, N)$, the procedure of the delete-1 cross-validation approach estimate of p_0 is given by \hat{p} .

The investigation on the limiting behavior of the procedure for $p > p_0$ is carried on as follows:

Denote $\Gamma^{(p)} = \sigma^2 I_{p+1} + \Omega D D^* \Omega^* = (\gamma_{m\ell}^{(p)})$, where $D = \text{diag}(\alpha_1, \dots, \alpha_{p_0})$ and

$$\Omega = \begin{bmatrix} 1 & \dots & 1 \\ e^{i\omega_1} & \dots & e^{i\omega_{p_0}} \\ \dots & \dots & \dots \\ e^{ip\omega_1} & \dots & e^{ip\omega_{p_0}} \end{bmatrix}. \quad (2.3)$$

Write $\hat{\Gamma}_p = (N - p)^{-1} Y^* Y$. By Lemma 3.2 of [1], it follows that

$$\hat{\Gamma}_p = \Gamma^{(p)} + O(\sqrt{\log \log(N - p)/(N - p)}), \quad \text{a.s.} \quad (2.4)$$

$$\hat{\Gamma}_{(-n,p)} = \Gamma^{(p)} + O(\sqrt{\log \log r_{(-n,p)}/r_{(-n,p)}}), \quad \text{a.s.} \quad (2.5)$$

Following the proof of Lemma 3.2 of [1], it can be shown that

$$\hat{\Gamma}_p - \hat{\Gamma}_{(-n,p)} = O(1/N). \quad (2.6)$$

Let $\lambda_1^{(p)} \geq \dots \geq \lambda_{p+1}^{(p)}$ be the eigenvalues of $\Gamma^{(p)}$, $\hat{\lambda}_1^{(p)} \geq \dots \geq \hat{\lambda}_{p+1}^{(p)}$ be the eigenvalues of $\hat{\Gamma}_p$, and $\hat{\lambda}_1^{(-n,p)} \geq \dots \geq \hat{\lambda}_{p+1}^{(-n,p)}$ be the eigenvalues of $\hat{\Gamma}_{(-n,p)}$. By [9],

$$\sum_{j=1}^{p+1} (\hat{\lambda}_j^{(p)} - \lambda_j^{(p)})^2 \leq \text{trace}(\hat{\Gamma}_p - \Gamma^{(p)})^2, \quad (2.7)$$

$$\sum_{j=1}^{p+1} (\hat{\lambda}_j^{(-n,p)} - \lambda_j^{(p)})^2 \leq \text{trace}(\hat{\Gamma}_{(-n,p)} - \Gamma^{(p)})^2. \quad (2.8)$$

Let $\hat{\mathbf{b}}^{(p)} = (\hat{b}_0^{(p)}, \dots, \hat{b}_p^{(p)})'$ be an unit eigenvector of $\hat{\Gamma}_p$ corresponding to $\hat{\lambda}_{p+1}^{(p)}$. By (2.4), (2.5), (2.7) and (2.8), with appropriate choices of $\hat{\mathbf{b}}^{(-n,p)}$ and $\hat{\mathbf{b}}^{(p)}$, we have that for $p > p_0$,

$$\hat{\mathbf{b}}^{(-n,p)} \rightarrow (\mathbf{b}' \ \mathbf{0}')', \quad \hat{\mathbf{b}}^{(p)} \rightarrow (\mathbf{b}' \ \mathbf{0}')',$$

where \mathbf{b} is a unit eigenvector of $\Gamma^{(p_0)}$ corresponding to its smallest eigenvalue σ^2 and $\mathbf{0}$ is a vector of $p - p_0$ zeros.

Denote the solutions of $\sum_{j=0}^p \hat{b}_j^{(p)} z^j = 0$ by $\hat{\rho}_j^{(p)} e^{-i\hat{\omega}_j^{(p)}}$, $j = 1, \dots, p$, where $\hat{\rho}_j^{(p)} > 0$, $\hat{\omega}_j^{(p)} \in [0, 2\pi)$, $j = 1, \dots, p$. When $p > p_0$ and with an appropriate ordering, it is easy to see that

$$\hat{\rho}_j^{(-n,p)} \rightarrow 1, \quad \hat{\rho}_j^{(p)} \rightarrow 1, \quad \hat{\omega}_j^{(-n,p)} \rightarrow \omega_j, \quad \hat{\omega}_j^{(p)} \rightarrow \omega_j, \quad (2.9)$$

for $j = 1, \dots, p_0$ and

$$\hat{\rho}_k^{(-n,p)} \rightarrow 0, \quad \hat{\rho}_k^{(p)} \rightarrow 0, \quad (2.10)$$

for $k = p_0 + 1, \dots, p$.

Let $q_{(-n,p)} = \hat{\mathbf{x}}_{(n,p)}^* (\hat{X}_{(-n,p)}^* \hat{X}_{(-n,p)})^{-1} \hat{\mathbf{x}}_{(n,p)}$ and $\hat{\boldsymbol{\alpha}}_{(-n,p)} = (\hat{X}_{(-n,p)}^* \hat{X}_{(-n,p)})^{-1} \hat{X}_{(-n,p)}^* \mathbf{y}_{(N)}$ for $n = 1, \dots, N$. From the fact that $(\hat{X}_{(-n,p)}^* \hat{X}_{(-n,p)})^{-1} = O(1/N)$, it is easy to see that

$$\lim_{N \rightarrow \infty} \max_{n \leq N} q_{(-n,p)} = 0, \quad \text{for any } p.$$

Since the square error is

$$\begin{aligned} & \|y(n) - \hat{\mathbf{x}}_{(n,p)} \hat{\boldsymbol{\alpha}}_{(-n,p)}\|^2 \\ &= \|(1, -\hat{\mathbf{x}}_{(n,p)} (\hat{X}_{(-n,p)}^* \hat{X}_{(-n,p)})^{-1} \hat{X}_{(-n,p)}^*) (y(n), \mathbf{y}'_{(-n,N)})'\|^2 \\ &= \|(1 - q_{(-n,p)})^{-1} [y(n) - \hat{\mathbf{x}}_{(n,p)}^* \hat{\boldsymbol{\alpha}}_{(-n,p)}]\|^2, \end{aligned}$$

it follows that

$$CR(p, N) = (1/N) \sum_{n=1}^N \|(1 - q_{(-n,p)})^{-1} [y(n) - \hat{\mathbf{x}}_{(n,p)} \hat{\boldsymbol{\alpha}}_{(-n,p)}]\|^2.$$

$$\text{As } (1 - q_{(-n,p)})^{-2} = 1 + 2q_{(-n,p)} + O(q_{(-n,p)}^2),$$

$$CR(p, N) = (1/N) \sum_{n=1}^N \|d_{(-n,p)}\|^2 + (1/N) \sum_{n=1}^N [2q_{(-n,p)} + O(q_{(-n,p)}^2)] \|d_{(-n,p)}\|^2,$$

where $d_{(-n,p)} = y(n) - \hat{\mathbf{x}}_{(n,p)}^* \hat{\boldsymbol{\alpha}}_{(-n,p)}$. By (2.6) and the fact that the roots of a polynomial are continuous functions of its coefficients, it follows that

$$|\hat{X}_{(-n,p)} (\hat{X}_{(-n,p)}^* \hat{X}_{(-n,p)})^{-1} \hat{X}_{(-n,p)}^* - \hat{X}_{(p)} (\hat{X}_{(p)}^* \hat{X}_{(p)})^{-1} \hat{X}_{(p)}^*| = O(1/N).$$

Therefore, for $p > p_0$,

$$\begin{aligned} CR(p, N) - CR(p_0, N) &= (1/N) \mathbf{y}_{(N)}^* [P_{\hat{X}_{(p)}} - P_{\hat{X}_{(p_0)}}] \mathbf{y}_{(N)} \\ &+ (1/N) \tilde{\mathbf{w}}_{(p, N)}^* D_{(p)} \tilde{\mathbf{w}}_{(p, N)} - (1/N) \tilde{\mathbf{w}}_{(p_0, N)}^* D_{(p_0)} \tilde{\mathbf{w}}_{(p_0, N)} + o_p(1/N), \end{aligned}$$

where $\tilde{\mathbf{w}}_{(p, N)} = [I - P_{\hat{X}_{(p)}}] \mathbf{y}_{(N)}$ with $P_A = A(A^*A)^{-1}A^*$, $D_{(p)} = \text{diag}(q_{(1, p)}, \dots, q_{(N, p)})$, $q_{(n, p)} = \boldsymbol{\xi}_{(n, p)}^* (\hat{X}_{(p)}^* \hat{X}_{(p)})^{-1} \boldsymbol{\xi}_{(n, p)}$, and $\boldsymbol{\xi}_{(n, p)} = (e^{in\hat{\omega}_1^{(p)}}, \dots, e^{in\hat{\omega}_p^{(p)}})'$, for $n = 1, \dots, N$. Hence, by (2.9) and (2.10),

$$Pr(CR(p, N) - CR(p_0, N) < 0) \geq c + o(1), \quad \text{for a constant } c > 0,$$

which means that the procedure does not select the true number of super imposed exponential signals with probability one.

3. THE DELETE- k CROSS-VALIDATION

Let τ be an index set and $A_{(-\tau, N, p)}$ be the matrix obtained from the matrix Y with the rows having the observations $y(n)$, $n \in \tau$ removed, $\theta_{(-\tau, N, p)} = \{n : (y(n + p), \dots, y(n)) \text{ is a row of } A_{(-\tau, N, p)}\}$, and $r_{(-\tau, N, p)}$ be the number of rows of $A_{(-\tau, N, p)}$. For convenience, we omit N in the above notations.

Denote $\hat{\Gamma}_{(-\tau, p)} = r_{(-\tau, p)}^{-1} A_{(-\tau, p)}^* A_{(-\tau, p)} = (\hat{\gamma}_{ml}^{(-\tau, p)})$, and let $\hat{\lambda}_1^{(-\tau, p)} \geq \dots \geq \hat{\lambda}_{p+1}^{(-\tau, p)}$ be the eigenvalues of $\hat{\Gamma}_{(-\tau, p)}$ and $\hat{\mathbf{b}}^{(-\tau, p)} = (\hat{b}_0^{(-\tau, p)}, \dots, \hat{b}_p^{(-\tau, p)})'$ be a unit eigenvector of $\hat{\Gamma}_{(-\tau, p)}$ corresponding to $\hat{\lambda}_{p+1}^{(-\tau, p)}$. Denote the solutions of $\sum_{j=0}^p \hat{b}_j^{(-\tau, p)} z^j = 0$ by $\hat{\rho}_j^{(-\tau, p)} e^{-i\hat{\omega}_j^{(-\tau, p)}}$, $j = 1, \dots, p$, where $\hat{\rho}_j^{(-\tau, p)} > 0$. Write

$$\hat{X}_{(-\tau, p)} = \begin{bmatrix} e^{i\hat{\omega}_1^{(-\tau, p)}} & \dots & e^{i\hat{\omega}_p^{(-\tau, p)}} \\ \dots & \dots & \dots \\ e^{iN\hat{\omega}_1^{(-\tau, p)}} & \dots & e^{iN\hat{\omega}_p^{(-\tau, p)}} \end{bmatrix},$$

$\tilde{X}_{(-\tau, p)}$ is a matrix with the n th row removed from $\hat{X}_{(-\tau, p)}$ for all $n \in \tau$, and $\mathbf{y}_{(-\tau, N)}$ is a random vector with the n th observation removed from $\mathbf{y}(N)$ for all $n \in \tau$. Define

$$\tilde{\alpha}_{(-\tau, p)} = (\tilde{X}_{(-\tau, p)}^* \tilde{X}_{(-\tau, p)})^{-1} \tilde{X}_{(-\tau, p)}^* \mathbf{y}_{(-\tau, N)}, \quad (3.1)$$

and for $\tau = \{j_1 < \dots < j_\ell\}$, write

$$\hat{X}_{(\tau,p)} = \begin{bmatrix} e^{ij_1\hat{\omega}_1^{(-\tau,p)}} & \dots & e^{ij_1\hat{\omega}_p^{(-\tau,p)}} \\ \dots & \dots & \dots \\ e^{ij_\ell\hat{\omega}_1^{(-\tau,p)}} & \dots & e^{ij_\ell\hat{\omega}_p^{(-\tau,p)}} \end{bmatrix},$$

and $\mathbf{y}_{(\tau,N)}$ is a random vector with the n th observation removed from $\mathbf{y}(N)$ for all $n \notin \tau$. Suppose that \mathcal{B} is a collection of $|\mathcal{B}|$ subsets of $\{1, \dots, N\}$ that have size k . Hereon, $|Q|$ denotes the number of elements in the set Q . Let

$$CR(p, k, \mathcal{B}, N) = \frac{1}{k|\mathcal{B}|} \sum_{\tau \in \mathcal{B}} \|\mathbf{y}_{(\tau,N)} - \hat{X}_{(\tau,p)} \tilde{\boldsymbol{\alpha}}_{(-\tau,p)}\|^2,$$

for $p = 0, \dots, P$. If $CR(\hat{p}, k, \mathcal{B}, N) = \min_{1 \leq p \leq P} CR(p, k, \mathcal{B}, N)$, the procedure based on the delete- k cross-validation approach uses \hat{p} to estimate p_0 .

\mathcal{B} can be chosen as the collection of all possible subsets of $\{1, \dots, N\}$ that have size k . If this is the case, then the procedure above is the standard delete- k cross-validation and it can be observed that the probability of overestimating p_0 is greater than $c > 0$ for large N by the same argument as in Section 2. [7] suggested to select \mathcal{B} with $|\mathcal{B}| = O(N)$ and the following conditions satisfied: (a) every n , $1 \leq n \leq N$, appears in the same number of subsets in \mathcal{B} ; and (b) every pair (n, m) , $1 \leq n < m \leq N$, appears in the same number of subsets in \mathcal{B} . Since the “balanced” collection \mathcal{B} is hard to obtain, [7] provided a simple and easy method by Monte Carlo, it proceeds as follows: randomly draw (with or without replacement) a collection \mathcal{G} of $|\mathcal{G}|$ subsets of $\{1, \dots, N\}$ that have size k with $|\mathcal{G}|^{-1}(N - k)^{-2}N^2 = o(1)$ and determine the number of super imposed exponential signals by minimizing

$$CR(p, k, \mathcal{G}, N) = \frac{1}{k|\mathcal{G}|} \sum_{\tau \in \mathcal{G}} \|\mathbf{y}_{(\tau,N)} - \hat{X}_{(\tau,p)} \tilde{\boldsymbol{\alpha}}_{(-\tau,p)}\|^2.$$

Their performances are studied by Monte Carlo simulation in the next section.

4. SIMULATION

The simulations for procedures in previous sections are carried out for the following two super imposed signal model:

$$y(n) = \alpha_1 \exp(in2\pi f_1) + \alpha_2 \exp(in2\pi f_2) + w(n), \quad n = 1, \dots, N,$$

where $f_1 = 1/4$, $f_2 = 1/24$ and $\alpha_1 = 1/\sqrt{2} + i/\sqrt{2}$, $\alpha_2 = 1$. The signal noise ratio (SNR) is defined as

$$SNR = 10 \log_{10} \frac{1}{p_0} \sum_{j=1}^{p_0} \frac{|\alpha_j|^2}{\sigma^2}.$$

In the simulation the number of possible signals P is given as 5 and the SNR is 42. For a given index set τ with k elements the matrix $A_{(-\tau, p)}$ may be constructed in the worst case scenario when

$$N - p > (k + 1)(p + 1). \quad (4.1)$$

This places a limitation on the rate of increase of k with that of N . In the all subset case (the number of " k -out" subsets is ${}_N C_k$) the inequality must be satisfied. For the Monte Carlo method this inequality may be overlooked at times due to the randomness of the selection of the k elements.

There are two types of simulations, one with the incomplete block design and the other using the Monte Carlo method (without replacement). The blocks were found by using the finite geometry constructed in the form of $N = q^2 + q + 1$ where q is a prime. This gives $q + 1$ elements in a block. Each pair of elements appear in only one block and each element appears in exactly $q + 1$ blocks. In this case one has the $q + 1$ deletions of elements and $|\mathcal{B}| = N$. Clearly with this kind of block design the inequality (4.1) is satisfied. For the Monte Carlo procedure, N , k and \mathcal{G} are chosen so that $|\mathcal{G}|^{-1}(N - k)^{-2}N^2$ decreases as N increases. For each size N and subset of k deletions the number of repetitions is 200 in both simulations.

Table 1

Sample Size	57	133	183	307	381	553	871	993	1407
No. of Deletions	8	12	14	18	20	24	30	32	38
Freq. of Right Detection	.12	.10	.14	.14	.13	.11	.07	.09	.10

The block design method of specifying the subset for deletions appears to be performing poorly in the simulations. Its frequency of correct detection is given in Table 1 along with the size N and number of deletion k . On the other hand the Monte Carlo procedure seems to give better results if the number of deletions is chosen to be large. The frequency of right detection for this procedure is plotted in Figure 1 where the ratio of k/N is also shown. The curves clearly show that when the ratio increases so does the frequency. This simulation result suggests that to obtain consistency of estimates, the ratio of k/N needs to increase with N and this would agree with the linear model selection problem considered in [7].

5. CONCLUSION

The intuitively appealing cross-validation technique for estimating the number of exponential signals is considered. For fixed number of k deletions ($k \geq 1$), it is proved that there is a positive probability to overestimate. This result agrees with that of the linear regression model studied by [7]. In [7], it is shown that in order to obtain a consistent linear model selection criterion one needs to vary the deletions in such a way that the number of deletions should increase with the sample size. Contrary to the linear model, there is a restriction on the way that the deletions are chosen here. To get good estimate of the number of signals, the number of deletions should be large and yet the matrix $A_{(-\tau, p)}$ may be constructed. This tendency was supported by the Monte-Carlo simulations.

References

- [1] Z. D. Bai, C. R. Rao, Y. Wu, M. M. Zen and L. C. Zhao, "Asymptotic properties of the simultaneous estimation of the number of signals and frequencies of multiple

sinusoids when some observations are missing: I. Asymptotics," *Proc. Natl. Acad. Sci. USA*. To appear.

- [2] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd International symposium on Information Theory*, B. N. Petrov and F. Csáki Eds. Akadémiai Kiado, Budapest, pp. 267-281, 1973.
- [3] D. Kundu and R. Kundu, "Consistent estimates of super imposed exponential signals when some observations are missing," *J. of Statistical Planning and inference*, vol. 44, pp. 205-218, 1995.
- [4] C. R. Rao, "Some results in signal detection," in *Decision Theory and Related Topics, IV*, S. S. Gupta and J. O. Berger Eds. vol. 2, Springer, New York, pp. 319-332, 1988.
- [5] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 468-478, 1978.
- [6] ———, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.
- [7] J. Shao, "Linear model selection by cross-validation," *J. Amer. Statist. Assoc.* vol. 88, pp. 486-494, 1993.
- [8] ———, "An asymptotic theory for linear model selection," *Statistica Sinica*, vol. 7, pp. 221-264, 1997.
- [9] J. von Neumann, "it Some matrix-inequalities and metrization of metric space," *Tomsk University*, Rev. 1, pp. 286-300, 1937.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB NO. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 25, 1999		3. REPORT TYPE AND DATES COVERED Technical - 99-13
4. TITLE AND SUBTITLE A Note of Estimating the Number of Super Imposed Exponential Signals by the Cross-Validation Approach			5. FUNDING NUMBERS DAHH04-96-1-0082	
6. AUTHOR(S) Y. Wu, K.W. Tam, F. Li and M.M. Zen				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Multivariate Analysis 417 Thomas Building Department of Statistics Penn State University University Park, PA 16802			8. PERFORMING ORGANIZATION REPORT NUMBER 99-13	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12 b. DISTRIBUTION CODE	
13. ABSTRACT In this paper, a procedure based on the delete-1 cross-validation is given for estimating the number of super imposed exponential signals, its limiting behavior is explored and it is shown that the probability of overestimating the true number of signals is larger than a positive constant for sample size large enough. Also a general procedure based on the cross-validation is presented when the deletion precedes according to a collection of subsets of indices. The result is similar to the delete-1 cross-validation if the number of deletion is fixed. The simulation results are provided for the performance of the procedure when the collections of subsets of indices are chosen as those suggested by [7] in a linear model selection problem.				
14. SUBJECT TERMS Exponential Signals, Cross Validation, Model Selection, Estimation of number of signals			15. NUMBER OF PAGES 10	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	